

**IN THE SPECIFICATION:**

**Please replace pages 10-14 ("BRIEF DESCRIPTION OF THE FIGURES") with the attached substitute pages 10-14.**

**Please rewrite the paragraph beginning at page 45, line 2 as follows:**

--Sequence of the pGEX and *mcg7* (underlined) junction:

pGEX-1      *mcg7* (695)

Sj26...GAA TTC GGC ACG AGC CGA CGG (SEQ ID NO: 20)

Additional amino acids:      Glu Phe Gly Thr Ser (SEQ ID NO: 111).--

**IN THE SEQUENCE LISTING:**

**Please replace the Sequence Listing of Record with the attached substitute Sequence Listing.**

**REMARKS**

Applicants respectfully submit that the specification has been amended to insert the sequence identifiers at pages 10-14 (Brief Description of the Figures) for those sequences disclosed in the figures. The specification has also been amended at page 45 to insert the sequence identifier for the amino acid sequence disclosed at line 2 of page 45. Applicants submit that the foregoing amendment does not introduce new matter.

Applicants have also amended the Sequence Listing of record. More specifically, after SEQ ID NO: 9 on page 74 and before SEQ ID NO: 10 on page 75, there are three sequences that are designated as SEQ ID NO: 7, SEQ ID NO: 8, and SEQ ID NO: 9. Applicants respectfully submit that these sequences were introduced as a result of a clerical error, and have

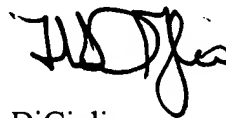
been deleted from the Sequence Listing as indicated in the attached marked-up copy of the Sequence Listing. In addition, Applicants have reformatted the sequences of SEQ ID NO: 1 to SEQ ID NO: 45 using the PatentIn Version 2.1. Furthermore, Applicants have added sequences designated as SEQ ID NOS: 46-111, which are disclosed in the Figures of the application as originally filed.

Applicants submit that the foregoing amendment to the Sequence Listing does not introduce new matter. Applicants submit herewith a substitute paper and initial computer readable copy of the Sequence Listing, along with a Statement Under 37 C.F.R. §1.821(f), stating that these copies are identical. A copy of the Notice to Comply is also enclosed.

Attached hereto is a marked-up version of the changes made to the specification and claims by the instant amendment. The attached page is captioned "**Version with Markings to Show Changes Made.**"

In view of the foregoing amendments and remarks, it is firmly believed that the subject application is in condition for allowance, which action is earnestly solicited.

Respectfully submitted,



Frank S. DiGiglio  
Registration No. 31,346

SCULLY, SCOTT, MURPHY & PRESSER  
400 Garden City Plaza  
Garden City, New York 11530  
(516) 742-4343

FSD/XZ:ab

Encls.: Substitute copy of pages 10-14 of the Specification  
Version with Markings to Show Changes Made  
Marked-up copy of the Sequence Listing

Serial No: 09/424,458



**VERSION WITH MARKINGS TO SHOW CHANGES MADE**

**IN THE SPECIFICATION:**

**Please amend the paragraph beginning at page 45, line 2, as follows:**

--Sequence of the pGEX and *mcg7* (underlined) junction:

pGEX-1      *mcg7* (695)

Sj26...GAA TTC GGC ACG AGC CGA CGG ({SEQ ID NO: 20})

Additional amino acids:      Glu Phe Gly Thr Ser (SEQ ID NO: 111).--

**Please amend pages 10-14 ("BRIEF DESCRIPTION OF THE FIGURES") as follows:**

--BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence ({SEQ ID NO: 2}) and corresponding amino acid sequence ({SEQ ID NO: 3}) of *mcg4*.

Figure 2 is a representation of the alignment of the human MCG4 amino acid sequence (amino acids: 1-60, SEQ ID NO: 46) (Query) with a translation of a partial murine expressed sequence tag (EST) (SEQ ID NO: 47) (Subject).

Figure 3 is a representation of the alignment of the human MCG4 amino acid sequence (amino acids 7-109; SEQ ID NO: 48) (Query) with a translation of a partial nematode EST (SEQ ID NO: 49) (subject).

Figure 4 is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn [represent] represents zinc atoms.

Figure 5 is a representation of sensitive sequence homology search of related cysteine-

containing motifs in another *Caenorhabditis elegans* protein. Query: amino acids 56-100 (SEQ ID NO: 50). Subject: SEQ ID NO: 51.

Figure 6 is a representation showing that a related cysteine containing motif is present in the GATA-binding transcription factor from *Saccharomyces pombe*. Queries: amino acids 35-58 (SEQ ID NO: 52); amino acids 162-227 (SEQ ID NO: 54). Subjects: amino acids 175-198 (SEQ ID NO: 53); amino acids 441-506 (SEQ ID NO: 55).

Figure 7 is a Northern blot showing expression of *mcg4* in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15 µg total RNA derived from various human cancer cells lines. Lanes 1-5, respectively, contain RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

Figure 8 is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788. Queries: nucleotides 446-704 (SEQ ID NO: 56); nucleotides 398-452 (SEQ ID NO: 58); nucleotides 767-810 (SEQ ID NO: 60); nucleotides 731-765 (SEQ ID NO: 62); nucleotides 701-732 (SEQ ID NO: 64); nucleotides 498-687 (SEQ ID NO: 66); nucleotides 398-495 (SEQ ID NO: 68); nucleotides 702-761 (SEQ ID NO: 70). Subjects: nucleotides 49-307 (SEQ ID NO: 57); nucleotides 2-56 (SEQ ID NO: 59); nucleotides 373-416 (SEQ ID NO: 61); nucleotides 336-370 (SEQ ID NO: 63); nucleotides 305-336 (SEQ ID NO: 65); nucleotides 103-292 (SEQ ID NO: 67); nucleotides 2-99 (SEQ ID NO: 69); nucleotides 309-368 (SEQ ID NO: 1).

Figure 9 is a representative of the partial nucleotide sequence alignment between a human (W32939) (SEQ ID NO: 72) and mouse (AA242159) (SEQ ID NO: 73) *mcg4*-like EST in the putative 5' UTR of the *mcg4* cDNA. The putative initiation codon is underlined and the region

upstream represents 5' UTR.

Figure 10 is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

Figure 11 is a diagrammatic representation of the domains of MCG4

zinc finger consensus:  $CX_2HX_4CX_2CX_4HX_2CX_{17}CX_2CX_{18}HX_2CX_{18}CX_2C$  (SEQ ID NO: 74);

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged;

basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged;

leucine zipper domain consensus:  $LX_6LX_6RX_6LX_6L$  (SEQ ID NO: 75);

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface to an alpha helix domain: (aa261)  $LX_6LXLX_6LXLX_6L$  (aa286) (SEQ ID NO: 76).

Figure 12 is a representation showing similarity of MCG7 WITH GEFs of various organisms.

Figure 13(a) is a representation of the nucleotide sequence (SEQ ID NO: 4) and corresponding amino acid sequence (SEQ ID NO: 5) of *mcg7*. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

Figure 13(b) is a representation of the partial nucleotide sequence (SEQ ID NO: 6) and corresponding amino acid sequence (SEQ ID NO: 7) OF *mcg7* but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon of nucleotides 183-288.

Figure 14 is a representation showing a comparison between MCG7 (SEQ ID NO: 7) and

a homologue from *Caenorhabditis elegans* (SEQ ID NO: 77) using the BESTFIT algorithm. In the figure, the following sequences are underlined:

**EF-Hand=PROSITE DATABASE NO. PD0C00018**

la nematode DVDEEDEVEDIEF ({SEQ ID NO: 10})  
 lb human DVDGDGHISQEEF ({SEQ ID NO: 11})  
 nematode DHDRDGFISQEEF ({SEQ ID NO: 12})  
 lc human DQNQDGCISREEM ({SEQ ID NO: 13})  
 nematode DVDMDGQISKDEL ({SEQ ID NO: 14})

**GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B**

2 human HFVHVAEKLLQLQNFNLTMAVVGGLSHSSISRLKETH ({SEQ ID NO: 15})  
 nematode KFVHVAKHLRKINNFNLTMSVVGGLSHSSVARLAKTY ({SEQ ID NO: 16})

**DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379**

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC  
 ({SEQ ID NO: 17})  
 nematode HNFHETTFLTPPTCNHCNKLLWGILRQGFCKDCGLAVHSCCKSNAVAEC  
 ({SEQ ID NO: 18})

Figure 15 is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine *mcg7* cDNA (GenBank Acc.No. W71787 and AA237373). The putative initiation codon is underlined. The murine sequence (SEQ ID NO: 78) represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

Figure 16 is a representation of further 5' nucleotide (SEQ ID NO: 79) and corresponding amino acid sequence (SEQ ID NO: 80) human *mcg7*. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Figure 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.

Figure 17 is a graphical representation of a GDP release assay. □Experiment #1 (mean of duplicate). ◇Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 min.

Estimated reaction constants:

$K_m = 2.1 \mu M$ ,  $V_{max} = 37 \text{ pMol/6min/36pMol}$  [Expt.#1]

$K_m = 1.5 \mu M$ ,  $V_{max} = 30.3 \text{ pMol/6min/36pMol}$  [Expt.#2]

Figure 18 depicts various recombinant plasmids containing partial or full-length *mcg7*.

Figure 19 is a representation of the nucleotide sequence ({SEQ ID NO: 8}) and corresponding amino acid sequence ({SEQ ID NO: 9}) of *mcg18*.

Figure 20 is representation showing that MCG18 (amino acids 35-96, SEQ ID NO: 81) (Query) has partial homology to *E. coli* DnaJ (amino acids 6-65, SEQ ID NO: 82) (Subject).

Figure 21 is a representation showing that MCG18 has homotology to two *Caenorhabditis elegans* proteins. Queries: amino acids 28-64 (SEQ ID NO: 83); amino acids 71-102 (SEQ ID NO: 85); amino acids 162-203 (SEQ ID NO: 87); amino acids 35-75 (SEQ ID NO: 89); amino acids 75-96 (SEQ ID NO:91); amino acids 141-184 (SEQ ID NO: 93). Subjects: amino acids 22-58 (SEQ ID NO:84); amino acids 64-95 (SEQ ID NO: 86); amino acids 158-199 (SEQ ID NO: 88); amino acids 19-50 (SEQ ID NO: 90); amino acids 71-92 (SEQ ID NO: 92); amino acids

**IN THE SEQUENCE LISTING:**

**Please amend the Sequence Listing of Record as indicated in the attached marked-up copy of the amended sequence listing.**



Marked-up Copy  
of the Sequence Listing.

- 57 -

Serial No.  
09/424,458

## SEQUENCE LISTING



## (i) GENERAL INFORMATION:

(i) APPLICANT: (OTHER THAN US): The Council of The Queensland Institute of Medical Research

(US ONLY): HAYWARD Nicholas, SILINS Ginters, GRIMMOND Sean,  
GARTSIDE Michael and HANCOCK, John

(ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 45

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DAVIES COLLISON CAVE

(B) STREET: 1 LITTLE COLLINS STREET

(C) CITY: MELBOURNE

(D) STATE: VICTORIA

(E) COUNTRY: AUSTRALIA

(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT INTERNATIONAL

(B) FILING DATE: 22-MAY-1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6973

(B) FILING DATE: 23-MAY-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6974

(B) FILING DATE: 23-MAY-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO6972

(B) FILING DATE: 23-MAY-1997

~~(C) CLASSIFICATION:~~~~(vii) PRIOR APPLICATION DATA:~~~~(A) APPLICATION NUMBER: PP1459~~~~(B) FILING DATE: 22-JAN-1998~~~~(C) CLASSIFICATION:~~~~(vii) PRIOR APPLICATION DATA:~~~~(A) APPLICATION NUMBER: PP1460~~~~(B) FILING DATE: 22-JAN-1998~~~~(C) CLASSIFICATION:~~~~(vii) PRIOR APPLICATION DATA:~~~~(A) APPLICATION NUMBER: PP1458~~~~(B) FILING DATE: 22-JAN-1998~~~~(C) CLASSIFICATION:~~~~(viii) ATTORNEY/AGENT INFORMATION:~~~~(A) NAME: HUGHES, DR E JOHN L~~~~(C) REFERENCE/DOCKET NUMBER: EJH/AF~~~~(ix) TELECOMMUNICATION INFORMATION:~~~~(A) TELEPHONE: +61 3 9254 2777~~~~(B) TELEFAX: +61 3 9254 2770~~~~(C) TELEX: AA-31787~~

<110> Hayward, Nicholas

Silinis, Ginters

Grimmond, Sean

Gartside, Michael

Hancock, John

<120> A Novel Gene and Uses Thereof

<130> 13198

<140> 09/424,458

<141> 7999-11-23

<160> 111

<170> PatentIn Ver. 2.1

- 59 -

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Xaa Xaa Cys Xaa Gly Xaa Gly  
 5

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1242 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 30..959

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCAGTAAACA CAGAGACTGG GGATCGATC ATG GGG CTT TGT AAG TGC CCC AAG	53
Met Gly Leu Cys Lys Cys Pro Lys	
1 5	
AGA AAG GTG ACC AAC CTG TTC TGC TTC GAA CAT CGG GTC AAC GTC TGC	101
Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys	
10 15 20	
GAG CAC TGC CTG GTA GCC AAT CAC GCC AAG TGC ATC GTC CAG TCC TAC	149
Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr	
25 30 35 40	
CTG CAA TGG CTC CAA GAT AGC GAC TAC AAC CCC AAT TGC CGC CTG TGC	197
Leu Gln Trp Leu Gln Asp Ser Asp Tyr Asn Pro Asn Cys Arg Leu Cys	
45 50 55	
AAC ATA CCC CTG GCC AGC CGA GAG ACG ACC CGC CTT GTC TGC TAT GAT	245
Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp	
60 65 70	
CTC TTT CAC TGG GCC TGC CTC AAT GAA CGT GCT GCC CAG CTA CCC CGA	293
Leu Phe His Trp Ala Cys Leu Asn Glu Arg Ala Ala Gln Leu Pro Arg	
75 80 85	
AAC ACG GCA CCT GCC GGC TAT CAG TGC CCC AGC TGC AAT GGC CCC ATC	341
Asn Thr Ala Pro Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ile	
90 95 100	
TTC CCC CCA ACC AAC CTG GCT GGC CCC GTG GCC TCC GCA CTG AGA GAG	389
Phe Pro Pro Thr Asn Leu Ala Gly Pro Val Ala Ser Ala Leu Arg Glu	
105 110 115 120	

- 60 -

AAG CTG GCC ACA GTC AAC TGG GCC CGG GCA GGA CTG GGC CTC CCT CTG Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu 125 130 135	437
ATC GAT GAG GTG GTG AGC CCA GAG CCC GAG CCC CTC AAC ACG TCT GAC Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp 140 145 150	485
TTC TCT GAC TGG TCT AGT TTT AAT GCC AGC AGT ACC CCT GGA CCA GAG Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu 155 160 165	533
GAG GTA GAC AGC GCC TCT GCT GCC CCA GCC TTC TAC AGC CGA GCC CCC Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro 170 175 180	581
CGG CCC CCA GCT TCC CCA GGC CGG CCC GAG CAG CAC ACA GTG ATC CAC Arg Pro Pro Ala Ser Pro Gly Arg Pro Glu Gln His Thr Val Ile His 185 190 195 200	629
ATG GGC AAT CCT GAG CCC TTG ACT CAC GCC CCT AGG AAG GTG TAT GAT Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp 205 210 215	677
ACG CGG GAT GAT GAC CGG ACA CCA GGC CTC CAT GGA GAC TGT GAC GAT Thr Arg Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp 220 225 230	725
GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTG CTA Asp Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu 235 240 245	773
AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 255 260	821
GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280	869
GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295	917
CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310	962
GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT	1022
AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA	1082
CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG	1142
GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC	1202
ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT	1242

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 310 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

- 61 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Leu Cys Lys Cys Pro Lys Arg Lys Val Thr Asn Leu Phe Cys  
 1 5 10 15  
 Phe Glu His Arg Val Asn Val Cys Glu His Cys Leu Val Ala Asn His  
 20 25 30  
 Ala Lys Cys Ile Val Gln Ser Tyr Leu Gln Trp Leu Gln Asp Ser Asp  
 35 40 45  
 Tyr Asn Pro Asn Cys Arg Leu Cys Asn Ile Pro Leu Ala Ser Arg Glu  
 50 55 60  
 Thr Thr Arg Leu Val Cys Tyr Asp Leu Phe His Trp Ala Cys Leu Asn  
 65 70 75 80  
 Glu Arg Ala Ala Gln Leu Pro Arg Asn Thr Ala Pro Ala Gly Tyr Gln  
 85 90 95  
 Cys Pro Ser Cys Asn Gly Pro Ile Phe Pro Pro Thr Asn Leu Ala Gly  
 100 105 110  
 Pro Val Ala Ser Ala Leu Arg Glu Lys Leu Ala Thr Val Asn Trp Ala  
 115 120 125  
 Arg Ala Gly Leu Gly Leu Pro Leu Ile Asp Glu Val Val Ser Pro Glu  
 130 135 140  
 Pro Glu Pro Leu Asn Thr Ser Asp Phe Ser Asp Trp Ser Ser Phe Asn  
 145 150 155 160  
 Ala Ser Ser Thr Pro Gly Pro Glu Glu Val Asp Ser Ala Ser Ala Ala  
 165 170 175  
 Pro Ala Phe Tyr Ser Arg Ala Pro Arg Pro Pro Ala Ser Pro Gly Arg  
 180 185 190  
 Pro Glu Gln His Thr Val Ile His Met Gly Asn Pro Glu Pro Leu Thr  
 195 200 205  
 His Ala Pro Arg Lys Val Tyr Asp Thr Arg Asp Asp Asp Arg Thr Pro  
 210 215 220  
 Gly Leu His Gly Asp Cys Asp Asp Asp Lys Tyr Arg Arg Arg Pro Ala  
 225 230 235 240  
 Leu Gly Trp Leu Ala Arg Leu Leu Arg Ser Arg Ala Gly Ser Arg Lys  
 245 250 255  
 Arg Pro Leu Thr Leu Leu Gln Arg Ala Gly Leu Leu Leu Leu Leu Gly  
 260 265 270  
 Leu Leu Gly Phe Leu Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg  
 275 280 285  
 Ala Ala Ala Asp Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His  
 290 295 300  
 Ile Arg Val Gly Pro Ser  
 305 310

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- 62 -

- (A) LENGTH: 2415 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 3..2186 2186

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT TCC Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser 1 5 10 15	47
CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG CTG TCC His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser 20 25 30	95
CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG CCT GCA GGA Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly 35 40 45	143
AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA ACG CAG CGC CTG Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu 50 55 60	191
TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA AGT GAA CAG CAC GTC Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val 65 70 75	239
CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT TCT GGG GTG GAC GAG CTG Gln Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu 80 85 90 95	287
GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly 100 105 110	335
CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC CTG GAC Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp 115 120 125	383
AAG GGC TGC ACG GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe 130 135 140	431
GAT GAC TCC GGG AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu 145 150 155	479
ATG ATG CAC CCC TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu 160 165 170 175	527
CTC CAC ATC TAC CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln 180 185 190	575
GTG AAA ACG TGC CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG Val Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala 195 200 205	623

- 63 -

GAG TTT GAC TTG AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG Glu Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys 210 215 220	671
GCT CTG CTA GAC CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC Ala Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp 225 230 235	719
ATA GAC AGC GTC CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG Ile Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg 240 245 250 255	767
AAC CCT GTG GGA CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC Asn Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His 260 265 270	815
CTG GAG CCC ATG GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg 275 280 285	863
TCC TTC TGC AAG ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT Ser Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His 290 295 300	911
GGC TGC ACT GTG GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC Gly Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe 305 310 315	959
AAC AGC GTC TCA CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA Asn Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr 320 325 330 335	1007
GCC CCG CAG CGG GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu 340 345 350	1055
AAG CTG CTA CAG CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG Lys Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly 355 360 365	1103
GGC CTG AGC CAC AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His 370 375 380	1151
GTT AGC CCT GAG ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val 385 390 395	1199
ACG GCG ACA GGC AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Leu Ala Ala Cys 400 405 410 415	1247
GTG GGC TTC CGC TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val 420 425 430	1295
GCC CTG CAG CTG GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg 435 440 445	1343
CTC AAC GGG GCC AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu 450 455 460	1391
GCC ATG GTG ACC AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu 465 470 475	1439

CTG AGC CTG CTC ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu 480 485 490 495	1487
CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro 500 505 510	1535
ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu 515 520 525	1583
GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val 530 535 540	1631
GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val 545 550 555	1679
GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly 560 565 570 575	1727
AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp 580 585 590	1775
GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser 595 600 605	1823
TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser 610 615 620	1871
AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu 625 630 635	1919
GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys 640 645 650 655	1967
CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala 660 665 670	2015
CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His 675 680 685	2063
AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg 690 695 700	2111
CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val 705 710 715	2159
GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TG GTTGGATC Glu Asp Gly Val Phe Asp Ile His Leu 720 725	2208
AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC	2268
GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC	2328



- 65 -

CAGGGCTGGT GTCCTAAGG TTGTACAGAC TCTTGTGAAT ATTTGTATTT TCCAGATGGA 2388  
ATAAAAAGGC CCGTGTAATT AACCTTC 2415

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 728 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His  
1 5 10 15  
Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro  
20 25 30  
Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg  
35 40 45  
Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys  
50 55 60  
Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln  
65 70 75 80  
Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly  
85 90 95  
Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro  
100 105 110  
Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys  
115 120 125  
Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp  
130 135 140  
Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met  
145 150 155 160  
Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu  
165 170 175  
His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val  
180 185 190  
Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu  
195 200 205  
Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala  
210 215 220  
Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile  
225 230 235 240  
Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn  
245 250 255  
Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu  
260 265 270

- 66 -

Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser  
 275 280 285  
 Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly  
 290 295 300  
 Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn  
 305 310 315 320  
 Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala  
 325 330 335  
 Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys  
 340 345 350  
 Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly  
 355 360 365  
 Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val  
 370 375 380  
 Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr  
 385 390 395 400  
 Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val  
 405 410 415  
 Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala  
 420 425 430  
 Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu  
 435 440 445  
 Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala  
 450 455 460  
 Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu  
 465 470 475 480  
 Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu  
 485 490 495  
 Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr  
 500 505 510  
 Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu  
 515 520 525  
 Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu  
 530 535 540  
 His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp  
 545 550 555 560  
 Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn  
 565 570 575  
 Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly  
 580 585 590  
 Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser  
 595 600 605  
 Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn  
 610 615 620  
 Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly

- 67 -

625	630	635	640
Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His			
	645	650	655
Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln			
	660	665	670
Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser			
	675	680	685
His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg			
	690	695	700
Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu			
	705	710	715
Asp Gly Val Phe Asp Ile His Leu			
	725		

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2309 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 254...2083 2080

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGATTTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG	60
CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCCGACCT CCACTAGGCC	120
TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAGC CCCATGGGAA	180
CGGGGTTTCGG TCCGAGCCCC GTGGGAGGCT CCCGGAGCGC AGCCTGGGCC CAGCCCCACC	240
CGCGCCGGCG GCC ATG GCA GGC ACC CTG GAC CTG GAC AAG GGC TGC ACG	289
Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr	
1 5 10	
GTG GAG GAG CTG CTC CGC GGG TGC ATC GAA GCC TTC GAT GAC TCC GGG	337
Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly	
15 20 25	
AAG GTG CGG GAC CCG CAG CTG GTG CGC ATG TTC CTC ATG ATG CAC CCC	385
Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro	
30 35 40	
TGG TAC ATC CCC TCC TCT CAG CTG GCG GCC AAG CTG CTC CAC ATC TAC	433
Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr	
45 50 55 60	
CAA CAA TCC CGG AAG GAC AAC TCC AAT TCC CTG CAG GTG AAA ACG TGC	481
Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys	
65 70 75	
CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG GAG TTT GAC TTG	529

- 68 -

His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu	
80 85 90	
AAC CCG GAG TTG GCT GAG CAG ATC AAG GAG CTG AAG GCT CTG CTA GAC	577
Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp	
95 100 105	
CAA GAA GGG AAC CGA CGG CAC AGC AGC CTA ATC GAC ATA GAC AGC GTC	625
Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val	
110 115 120	
CCT ACC TAC AAG TGG AAG CGG CAG GTG ACT CAG CGG AAC CCT GTG GGA	673
Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly	
125 130 135 140	
CAG AAA AAG CGC AAG ATG TCC CTG TTG TTT GAC CAC CTG GAG CCC ATG	721
Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met	
145 150 155	
GAG CTG GCG GAG CAT CTC ACC TAC TTG GAG TAT CGC TCC TTC TGC AAG	769
Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys	
160 165 170	
ATC CTG TTT CAG GAC TAT CAC AGT TTC GTG ACT CAT GGC TGC ACT GTG	817
Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val	
175 180 185	
GAC AAC CCC GTC CTG GAG CGG TTC ATC TCC CTC TTC AAC AGC GTC TCA	865
Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser	
190 195 200	
CAG TGG GTG CAG CTC ATG ATC CTC AGC AAA CCC ACA GCC CCG CAG CGG	913
Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg	
205 210 215 220	
GCC CTG GTC ATC ACA CAC TTT GTC CAC GTG GCG GAG AAG CTG CTA CAG	961
Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln	
225 230 235	
CTG CAG AAC TTC AAC ACG CTG ATG GCA GTG GTC GGG GGC CTG AGC CAC	1009
Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His	
240 245 250	
AGC TCC ATC TCC CGC CTC AAG GAG ACC CAC AGC CAC GTT AGC CCT GAG	1057
Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu	
255 260 265	
ACC ATC AAG CTC TGG GAG GGT CTC ACG GAA CTA GTG ACG GCG ACA GGC	1105
Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly	
270 275 280	
AAC TAT GGC AAC TAC CGG CGT CGG CTG GCA GCC TGT GTG GGC TTC CGC	1153
Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg	
285 290 295 300	
TTC CCG ATC CTG GGT GTG CAC CTC AAG GAC CTG GTG GCC CTG CAG CTG	1201
Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu	
305 310 315	
GCA CTG CCT GAC TGG CTG GAC CCA GCC CGG ACC CGG CTC AAC GGG GCC	1249
Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala	
320 325 330	
AAG ATG AAG CAG CTC TTT AGC ATC CTG GAG GAG CTG GCC ATG GTG ACC	1297
Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr	
335 340 345	

AGC CTG CGG CCA CCA GTA CAG GCC AAC CCC GAC CTG CTG AGC CTG CTC Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu 350 355 360	1345
ACG GTG TCT CTG GAT CAG TAT CAG ACG GAG GAT GAG CTG TAC CAG CTG Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu 365 370 375 380	1393
TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA ACC AGC CCC ACG Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr 385 390 395	1441
AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG GAG TGG ACC TCG Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser 400 405 410	1489
GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG GAG CAC ATC GAG Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu 415 420 425	1537
AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC GAT GGG GAT GGC Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly 430 435 440	1585
CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG AAC TTC CCT TAC His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr 445 450 455 460	1633
CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT GGC TGC ATC AGC Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser 465 470 475	1681
AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC TCT GTG TTG GGG Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly 480 485 490	1729
GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC AAC TCC TTG CGC Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg 495 500 505	1777
CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG GGC ATC TAC AAG Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys 510 515 520	1825
CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC CAC AAG CAG TGC Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys 525 530 535 540	1873
AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC CAG AGT GTG AGC Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser 545 550 555	1921
CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC AGC CAC CAT CAC Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser His His His 560 565 570	1969
CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG CGA GGC TCC AGG Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg 575 580 585	2017
CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG GAG GAT GGG GTG Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val 590 595 600	2065
TTT GAC ATC CAC TTG TAATAGATGC TGTGTTGGA TCAAGGACTC ATTCCTGCCT Phe Asp Ile His Leu 605 610	2120

- 70 -

TGGAGAAAAT ACTTCAACCA GAGCAGGGAG CCTGGGGGTG TCGGGGCAGG AGGCTGGGGA 2180  
 TGGGGGTGGG ATATGAGGGT GGCATGCAGC TGAGGGCAGG GCCAGGGCTG GTGTCCCTAA 2240  
 GGTGTACAG ACTCTTGTGA ATATTTGTAT TTTCCAGATG GAATAAAAAG GCCCGTGTA 2300  
 TTAACCTTC 2309

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 609 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 1 5 10 15  
 Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp  
 20 25 30  
 Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro  
 35 40 45  
 Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg  
 50 55 60  
 Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys His Leu Val Arg  
 65 70 75 80  
 Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu Asn Pro Glu Leu  
 85 90 95  
 Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp Gln Glu Gly Asn  
 100 105 110  
 Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val Pro Thr Tyr Lys  
 115 120 125  
 Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly Gln Lys Lys Arg  
 130 135 140  
 Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met Glu Leu Ala Glu  
 145 150 155 160  
 His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys Ile Leu Phe Gln  
 165 170 175  
 Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val Asp Asn Pro Val  
 180 185 190  
 Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser Gln Trp Val Gln  
 195 200 205  
 Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile  
 210 215 220  
 Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe  
 225 230 235 240  
 Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser  
 245 250 255

- 71 -

Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu  
 260 265 270  
 Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly Asn Tyr Gly Asn  
 275 280 285  
 Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg Phe Pro Ile Leu  
 290 295 300  
 Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu Ala Leu Pro Asp  
 305 310 315 320  
 Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln  
 325 330 335  
 Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro  
 340 345 350  
 Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu  
 355 360 365  
 Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu Ser Leu Gln Arg  
 370 375 380  
 Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr Ser Cys Thr Pro  
 385 390 395 400  
 Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser Ala Ala Lys Pro  
 405 410 415  
 Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu  
 420 425 430  
 Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln  
 435 440 445  
 Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe  
 450 455 460  
 Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
 465 470 475 480  
 Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly  
 485 490 495  
 Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys  
 500 505 510  
 Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys  
 515 520 525  
 Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu  
 530 535 540  
 Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser Leu Glu Gly Ser  
 545 550 555 560  
 Ala Pro Ser Pro Ser Pro Met His Ser His His His Arg Ala Phe Ser  
 565 570 575  
 Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg Pro Pro Glu Ile  
 580 585 590  
 Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val Phe Asp Ile His  
 595 600 605  
 Leu

- 72 -

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 832 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 11..733

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

GCCCGCCGCC ATG CCG CCC TTA CTG CCC CTG CGC CTG TGC CGG CTG TGG      49
Met Pro Pro Leu Leu Pro Leu Arg Leu Cys Arg Leu Trp
      1              5              10

CCC CGC AAC CCT CCC TCC CGG CTC CTC GGA GCG GCC GCC GGG CAG CGG      97
Pro Arg Asn Pro Pro Ser Arg Leu Leu Gly Ala Ala Ala Gly Gln Arg
      15              20              25

TCC AGA CCC AGT ACT TAT TAT GAA CTG TTG GGG GTG CAT CCT GGT GCC      145
Ser Arg Pro Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala
      30              35              40              45

AGC ACT GAG GAA GTT AAA CGA GCT TTC TTC TCC AAG TCC AAA GAG CTG      193
Ser Thr Glu Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu
      50              55              60

CAC CCA GAC CGG GAC CCT GGG AAC CCA AGC CTG CAC AGC CGC TTT GTG      241
His Pro Asp Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val
      65              70              75

GAG CTG AGC GAG GCA TAC CGT GTG CTC AGC CGT GAG CAG AGC CGC CGC      289
Glu Leu Ser Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg
      80              85              90

AGC TAT GAT GAC CAG CTC CGC TCA GGT AGT CCC CCA AAG TCT CCA CGA      337
Ser Tyr Asp Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg
      95              100              105

ACC ACA GTC CAT GAC AAG TCT GCC CAC CAA ACA CAC AGC TCC TGG ACA      385
Thr Thr Val His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr
      110              115              120              125

CCC CCC AAC GCA CAG TAC TGG TCC CAG TTT CAC AGC GTG AGG CCA CAG      433
Pro Pro Asn Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln
      130              135              140

GGG CCC CAG TTG AGG CAG CAG CAA CAC AAA CAA AAC AAA CAA GTG CTG      481
Gly Pro Gln Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu
      145              150              155

GGG TAC TGC CTC CTC CTC ATG CTG GCG GGC ATG GGC CTG CAC TAC ATT      529
Gly Tyr Cys Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile
      160              165              170

GCC TTC AGG AAG GTG AAG CAG ATG CAC CTT AAC TTC ATG GAT GAA AAG      577
Ala Phe Arg Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys
      175              180              185

```



- 73 -

GAT CGG ATC ATC ACA GCC TTC TAC AAC GAA GCC CGG GCA CGG GCC AGG 625  
 Asp Arg Ile Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg  
 190 195 200 205

GCC AAC AGA GGC ATC CTT CAG CAG GAG CGA CAA CGG CTA GGG CAG CGG 673  
 Ala Asn Arg Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg  
 210 215 220

CAG CCG CCA CCA TCC GAG CCA ACC CAA GGC CCC GAG ATC GTG CCC CGG 721  
 Gln Pro Pro Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg  
 225 230 235

GGC GCC GGC CCC TGA GGGGCTC ACCTGGATGG GGCCTGCAGT GCGTTCCCGC 773  
 Gly Ala Gly Pro \*  
 240

TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTCGCAG 832

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Pro Pro Leu Leu Pro Leu Arg Leu Cys Arg Leu Trp Pro Arg Asn  
 1 5 10 15

Pro Pro Ser Arg Leu Leu Gly Ala Ala Ala Gly Gln Arg Ser Arg Pro  
 20 25 30

Ser Thr Tyr Tyr Glu Leu Leu Gly Val His Pro Gly Ala Ser Thr Glu  
 35 40 45

Glu Val Lys Arg Ala Phe Phe Ser Lys Ser Lys Glu Leu His Pro Asp  
 50 55 60

Arg Asp Pro Gly Asn Pro Ser Leu His Ser Arg Phe Val Glu Leu Ser  
 65 70 75 80

Glu Ala Tyr Arg Val Leu Ser Arg Glu Gln Ser Arg Arg Ser Tyr Asp  
 85 90 95

Asp Gln Leu Arg Ser Gly Ser Pro Pro Lys Ser Pro Arg Thr Thr Val  
 100 105 110

His Asp Lys Ser Ala His Gln Thr His Ser Ser Trp Thr Pro Pro Asn  
 115 120 125

Ala Gln Tyr Trp Ser Gln Phe His Ser Val Arg Pro Gln Gly Pro Gln  
 130 135 140

Leu Arg Gln Gln Gln His Lys Gln Asn Lys Gln Val Leu Gly Tyr Cys  
 145 150 155 160

Leu Leu Leu Met Leu Ala Gly Met Gly Leu His Tyr Ile Ala Phe Arg  
 165 170 175

Lys Val Lys Gln Met His Leu Asn Phe Met Asp Glu Lys Asp Arg Ile  
 180 185 190

Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg Ala Asn Arg

- 74 -

195

200

205

Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro  
 210 215 220  
 Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly  
 225 230 235 240  
 Pro

SEQ ID Nos: 10-18 25-36

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 300 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 170..300

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAT ATATTCCCAT CTTGTCCTAG 60  
 CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCGCCACG CCCCCGACCT CCACTAGGCC 120  
 TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGTCCCGGG CCGGGTTAG CCC CAT 175  
 Pro His  
 1  
 GGG AAC GGG GTT CGG TCC GAG CCG GGT GGG AGG CTC CCG GAG CGC AGC 223  
 Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser  
 5 10 15  
 CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC 271  
 Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp  
 20 25 30  
 CTG GAC AAG GGC TGC ACG GTG GAG GAG CT 300  
 Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
 35 40

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 44 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 75 -

~~(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:~~

~~Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu  
1 5 10 15  
Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr  
20 25 30  
Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu  
35 40~~

~~(2) INFORMATION FOR SEQ ID NO:9:~~

- ~~(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear~~

~~(ii) MOLECULE TYPE: DNA~~~~(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:~~~~GGGATCCCCC TGGTC~~~~15~~~~(2) INFORMATION FOR SEQ ID NO:10:~~

- ~~(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear~~

~~(ii) MOLECULE TYPE: Peptide~~~~(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:~~

~~Asp Val Asp Glu Glu Asp Glu Val Glu Asp Ile Glu Phe  
1 5 10~~

~~(2) INFORMATION FOR SEQ ID NO:11:~~

- ~~(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear~~

~~(ii) MOLECULE TYPE: Peptide~~~~(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:~~

~~Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe  
1 5 10~~

~~(2) INFORMATION FOR SEQ ID NO:12:~~

- ~~(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids~~

- 76 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp His Asp Arg Asp Gly Phe Ile Ser Gln Glu Glu Phe  
 1                      5                      10

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met  
 1                      5                      10

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Val Asp Met Asp Gly Gln Ile Ser Lys Asp Glu Leu  
 1                      5                      10

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe Asn  
 1                      5                      10                      15

Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg  
 20                      25                      30

- 77 -

Leu Lys Glu Thr His  
35

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 37 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Phe Val His Val Ala Lys His Leu Arg Lys Ile Asn Asn Phe Asn  
 1 5 10 15  
 Thr Leu Met Ser Val Val Gly Gly Ile Thr His Ser Ser Val Ala Arg  
 20 25 30  
 Leu Ala Lys Thr Tyr  
 35

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys Arg His  
 1 5 10 15  
 Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg  
 20 25 30  
 Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val  
 35 40 45  
 Glu Cys  
 50

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

His Asn Phe His Glu Thr Thr Phe Leu Thr Pro Thr Thr Cys Asn His

- 78 -

1	5	10	15
Cys	Asn	Lys	Leu
Leu	Trp	Gly	Ile
Leu	Arg	Gln	Gly
Phe	Lys	Cys	Lys
20	25	30	
Asp	Cys	Gly	Leu
Ala	Val	His	Ser
Cys	Cys	Lys	Ser
Asn	Ala	Val	Ala
35	40	45	
Glu	Cys		
50			

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGGATCCCCC TGGTC

15

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTTCGGCA CGAGCCGACG G

21

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 78 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGGAGCAGA AGCTGATCTC CGAGGAGGAC CTGCCCCGGGG CAGCTGGATC CGCAGCCCAC  
 CCCGCGCCGG CGGCCATG

60

78

## (2) INFORMATION FOR SEQ ID NO:22:

- 79 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Pro Gly Ala Ala Gly  
1                      5                      10                      15  
Ser Ala Ala His Pro Ala Pro Ala Ala Met  
20                      25

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGATCCGCAG CCCACCCCGC GCCGGCGGCC ATG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met  
5                      10

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 80 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGACAAAGTG TGTGATGAAC C

21

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTCATCCTCC GTCTGATACT G

21

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GTAGATGTGG ATCAGCTTGG

20

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AGGTGGAGAA TGGTCAAGG

19

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTCATAGTCT GTCTCCTACT

20



- 81 -

## (2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ACATAGACAG CGTGCCTACC

20

## (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACAACCTTA GGGACACCAG

20

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TGCTGAGCCT GCTCACGGTG

20

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CAAGTGAACA GCACGTCC

18

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:

- 82 -

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GACTATCTCA AGGACCAGCT G

21

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGTTCGGTCC GAGCCCGG

18

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAGCGATAC TCCAAGTAGG T

21

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AGCGGGCCAG GCCCCTTC

18

(2) INFORMATION FOR SEQ ID NO:38:

- 83 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATCCTGGTC CAATGCGCTC

20

- (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCACTGAGGA AGTTAAACGA GC

22

- (2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCTCGTTTAA CTCCTCAGT GC

22

- (2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTCAGCTCC ACAAAGCGGC T

21

- (2) INFORMATION FOR SEQ ID NO:42:

- 84 -

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 19 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ACCAGCTCCG CTCAGGTAG

19

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TCCAGGAGCT GTGTGTTTGG

20

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 19 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CCAGTTTCAC AGCGTGAGG

19

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 19 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CAGCATGAGG AGGAGGCAG

19

*Please add SEQ IDs 46-110 as  
Set forth in the substitute Sequence Listing.*



## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence (SEQ ID NO: 2) and corresponding amino acid sequence (SEQ ID NO: 3) of *mcg4*.

Figure 2 is a representation of the alignment of the human MCG4 amino acid sequence (amino acids: 1-60, SEQ ID NO: 46) (Query) with a translation of a partial murine expressed sequence tag (EST) (SEQ ID NO: 47) (Subject).

Figure 3 is a representation of the alignment of the human MCG4 amino acid sequence (amino acids 7-109; SEQ ID NO: 48) (Query) with a translation of a partial nematode EST (SEQ ID NO: 49) (subject).

Figure 4 is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represents zinc atoms.

Figure 5 is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein. Query: amino acids 56-100 (SEQ ID NO: 50). Subject: SEQ ID NO: 51.

Figure 6 is a representation showing that a related cysteine containing motif is present in the GATA-binding transcription factor from *Saccharomyces pombe*. Queries: amino acids 35-58 (SEQ ID NO: 52); amino acids 162-227 (SEQ ID NO: 54). Subjects: amino acids 175-198 (SEQ ID NO: 53); amino acids 441-506 (SEQ ID NO: 55).

Figure 7 is a Northern blot showing expression of *mcg4* in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15 µg total RNA derived from various human cancer cells lines. Lanes 1-5, respectively, contain RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

Figure 8 is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788. Queries: nucleotides 446-704 (SEQ ID NO: 56); nucleotides 398-452 (SEQ ID NO: 58); nucleotides 767-810 (SEQ ID NO: 60); nucleotides 731-765 (SEQ ID NO: 62); nucleotides 701-732 (SEQ ID NO: 64); nucleotides 498-687 (SEQ ID NO: 66); nucleotides 398-495 (SEQ ID NO: 68); nucleotides 702-761 (SEQ ID NO: 70). Subjects: nucleotides 49-307 (SEQ ID NO: 57); nucleotides 2-56 (SEQ ID NO: 59); nucleotides 373-416 (SEQ ID NO: 61);

nucleotides 336-370 (SEQ ID NO: 63); nucleotides 305-336 (SEQ ID NO: 65); nucleotides 103-292 (SEQ ID NO: 67); nucleotides 2-99 (SEQ ID NO: 69); nucleotides 309-368 (SEQ ID NO: 1).

Figure 9 is a representative of the partial nucleotide sequence alignment between a human (W32939) (SEQ ID NO: 72) and mouse (AA242159) (SEQ ID NO: 73) *mcp4*-like EST in the putative 5' UTR of the *mcp4* cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

Figure 10 is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

Figure 11 is a diagrammatic representation of the domains of MCG4  
 zinc finger consensus: CX<sub>2</sub>HX<sub>4</sub>CX<sub>2</sub>CX<sub>4</sub>HX<sub>2</sub>CX<sub>17</sub>CX<sub>2</sub>CX<sub>18</sub>HX<sub>2</sub>CX<sub>18</sub>CX<sub>2</sub>C (SEQ ID NO: 74);  
 acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged;  
 basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged;  
 leucine zipper domain consensus: LX<sub>6</sub>LX<sub>6</sub>RX<sub>6</sub>LX<sub>6</sub>L (SEQ ID NO: 75);  
 alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface to an alpha helix domain: (aa261) LX<sub>6</sub>LXLX<sub>6</sub>LXLX<sub>6</sub>L (aa286) (SEQ ID NO: 76).

Figure 12 is a representation showing similarity of MCG7 WITH GEFs of various organisms.

Figure 13(a) is a representation of the nucleotide sequence (SEQ ID NO: 4) and corresponding amino acid sequence (SEQ ID NO: 5) of *mcp7*. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

Figure 13(b) is a representation of the partial nucleotide sequence (SEQ ID NO: 6) and corresponding amino acid sequence (SEQ ID NO: 7) OF *mcp7* but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon of nucleotides 183-288.

Figure 14 is a representation showing a comparison between MCG7 (SEQ ID NO: 7) and a homologue from *Caenorhabditis elegans* (SEQ ID NO: 77) using the BESTFIT algorithm. In the figure, the following sequences are underlined:

**EF-Hand=PROSITE DATABASE NO. PD0C00018**

la nematode DVDEEDEVEDIEF (SEQ ID NO: 10)  
 lb human DVDGDGHISQEEF (SEQ ID NO: 11)  
     nematode DHDRDGFISQEEF (SEQ ID NO: 12)  
 lc human DQNQDGCISREEM (SEQ ID NO: 13)  
     nematode DVDMDGQISKDEL (SEQ ID NO: 14)

**GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B**

2 human HFVHVAEKLLQLQNFNLTMAVVGGLSHSSISRLKETH (SEQ ID NO: 15)  
     nematode KFVHVAKHLRKINNFNLTMSVVGGLTHSSVARLAKTY (SEQ ID NO: 16)

**DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379**

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC  
     (SEQ ID NO: 17)  
     nematode HNFHETTFLPTTCNHCNKLLWGILRQGFCKKDCGLAVHSCCKSNAVAEC  
     (SEQ ID NO: 18)

Figure 15 is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine *mcg7* cDNA (GenBank Acc.No. W71787 and AA237373). The putative initiation codon is underlined. The murine sequence (SEQ ID NO: 78) represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

Figure 16 is a representation of further 5' nucleotide (SEQ ID NO: 79) and corresponding amino acid sequence (SEQ ID NO: 80) human *mcg7*. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Figure 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.

Figure 17 is a graphical representation of a GDP release assay. □Experiment #1 (mean of duplicate). ◇Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 min.

Estimated reaction constants:

$K_m = 2.1 \mu M$ ,  $V_{max} = 37 \text{ pMol/6min/36pMol [Expt.\#1]}$

$K_m = 1.5\mu\text{M}$ ,  $V_{\text{max}} = 30.3\text{pMol}/6\text{min}/36\text{pMol}[\text{Expt.}\#2]$

Figure 18 depicts various recombinant plasmids containing partial or full-length *mcg7*.

Figure 19 is a representation of the nucleotide sequence (SEQ ID NO: 8) and corresponding amino acid sequence (SEQ ID NO: 9) of *mcg18*.

Figure 20 is representation showing that MCG18 (amino acids 35-96, SEQ ID NO: 81] (Query) has partial homology to *E. coli* DnaJ (amino acids 6-65, SEQ ID NO: 82) (Subject).

Figure 21 is a representation showing that MCG18 has homology to two *Caenorhabditis elegans* proteins. Queries: amino acids 28-64 (SEQ ID NO: 83); amino acids 71-102 (SEQ ID NO: 85); amino acids 162-203 (SEQ ID NO: 87); amino acids 35-75 (SEQ ID NO: 89); amino acids 75-96 (SEQ ID NO: 91); amino acids 141-184 (SEQ ID NO: 93). Subjects: amino acids 22-58 (SEQ ID NO: 84); amino acids 64-95 (SEQ ID NO: 86); amino acids 158-199 (SEQ ID NO: 88); amino acids 19-50 (SEQ ID NO: 90); amino acids 71-92 (SEQ ID NO: 92); amino acids 145-188 (SEQ ID NO: 94).

Figure 22 is a representation showing that MCG18 has homology to a *Saccharomyces pombe* protein. Queries: amino acids 35-70 (SEQ ID NO: 95); amino acids 75-114 (SEQ ID NO: 97); amino acids 190-218 (SEQ ID NO: 99); amino acids 140-158 (SEQ ID NO: 101). Subjects: amino acids 9-44 (SEQ ID NO: 96); amino acids 50-89 (SEQ ID NO: 98); amino acids 149-177 (SEQ ID NO: 100); amino acids 44-62 (SEQ ID NO: 102).

Figure 23 is a representation showing homology of MCG18 to a *Drosophila virilis* protein. Queries: amino acids 26-95 (SEQ ID NO: 103). Subject: amino acids 72-142 (SEQ ID NO: 104).

Figure 24 is a representation showing homology of MCG18 to human DnaJ protein HDJ-2/HSDJ (SEQ ID NO: 105), HDJ-1/HSP40 (SEQ ID NO: 106) and HSP70 (SEQ ID NO: 107).

Figure 25 is a representation of the nucleotide (SEQ ID NO: 108) and corresponding amino acid sequence (SEQ ID NO: 109) of murine *mcg18*.

Figure 26 is a representation of homology between human and murine MCG18 (SEQ ID NO: 9 and SEQ ID NO: 109).

Figure 27 depicts nucleotide sequence (SEQ ID NO: 110) corresponding to the 5' untranslated region of human *mcg18*.



Figure 28 depicts a Northern blot showing expression of mcg 18 transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain 15µg RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.